

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-271774

(43)Date of publication of application : 20.10.1995

(51)Int.Cl. G06F 17/21
G06F 17/00

(21)Application number : 06-061534 (71)Applicant : SHARP CORP
(22)Date of filing : 30.03.1994 (72)Inventor : UEDA TORU
YUASA NATSUKI

(54) DOCUMENT INPUT DEVICE

(57)Abstract:

PURPOSE: To provide the document input device which has a function for lightening the burden of input by predicting and displaying a word appearing next by using information on the meaning that a word that is already inputted has.

CONSTITUTION: This document input device includes an input means 1 for inputting a document, a document buffer 2 which stores the inputted document, a meaning vector generating means 3 which generates meaning vectors of the document from the document stored in the document buffer 2, a word predicting means 4 which predicts following words from the meaning vectors of the document, a meaning dictionary 5 which is used to predict the words, a predicted word display means 6 which displays the predicted word, and a word selecting means 7 which selects one word out of the predicted words. Then the word predicting means 4 is equipped with a function which compares the similarity between the meaning vectors of the document and the meaning vectors of the words described in the meaning dictionary and selects the next word. Further, the word predicting means 4 has a function for narrowing down the predicted words by using information on the character which is inputted after predicting the words.



LEGAL STATUS

[Date of request for examination] 06.02.1998
[Date of sending the examiner's decision of rejection] 05.06.2001
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection] 2001-011514
[Date of requesting appeal against examiner's decision of rejection] 05.07.2001
[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-271774

(43)公開日 平成7年(1995)10月20日

(51)Int.Cl.⁶

識別記号

庁内整理番号

F I

技術表示箇所

G 0 6 F 17/21

17/00

9288-5L

8724-5L

G 0 6 F 15/ 20

5 6 4 E

F

審査請求 未請求 請求項の数 3 O L (全 7 頁)

(21)出願番号

特願平6-61534

(22)出願日

平成6年(1994)3月30日

(71)出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72)発明者 上田 徹

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(72)発明者 湯浅 夏樹

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

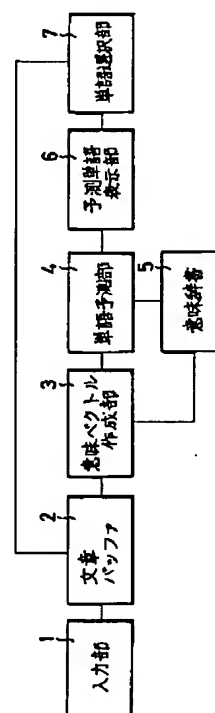
(74)代理人 弁理士 川口 義雄 (外1名)

(54)【発明の名称】 文章入力装置

(57)【要約】

【目的】 既に入力された言葉の持つ意味的な情報を用いて、次に出現する単語を予測／表示し、入力の手荷を軽減する機能を持った文章入力装置を提供する。

【構成】 文章を入力する入力手段と、入力された文章を格納する文章バッファと、文章バッファに蓄えられた文章から文章の意味ベクトルを作成する意味ベクトル作成手段と、文章の意味ベクトルから次の単語を予測する単語予測手段と、単語を予測するために使用する意味辞書と、予測された単語を表示する予測単語表示手段と、予測された複数の単語から一つを選択する単語選択手段とを含む。上記単語予測手段は、文章の意味ベクトルと意味辞書に記述された単語の意味ベクトルとの類似度を比較して次の単語を選別する機能を具備している。さらに、上記単語予測手段は、単語を予測した後に入力された文字の情報を用いて予測単語を絞り込む機能を具備している。



【特許請求の範囲】

【請求項1】 文章を入力する入力手段と、入力された文章を格納する文章バッファと、文章バッファに蓄えられた文章から文章の意味ベクトルを作成する意味ベクトル作成手段と、文章の意味ベクトルから次の単語を予測する単語予測手段と、単語を予測するために使用する意味辞書と、予測された単語を表示する予測単語表示手段と、予測された複数の単語から一つを選択する単語選択手段とを含むことを特徴とする文章入力装置。

【請求項2】 単語予測手段が、文章の意味ベクトルと意味辞書に記述された単語の意味ベクトルとの類似度を比較して次の単語を選別する機能を具備していることを特徴とする請求項1記載の文章入力装置。

【請求項3】 単語予測手段が、単語を予測した後に入力された文字の情報をを用いて予測単語を絞り込む機能を具備していることを特徴とする請求項1又は請求項2記載の文章入力装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、パソコン／電子手帳／ワープロなどの文章入力装置に関する。

【0002】

【従来の技術】 従来、ユーザが過去に入力したデータを活用して入力負荷を軽減する機能（学習機能）としては、かな漢字変換時に既に文章中で入力された単語を、変換候補の上位にするという方法がある。また、既に入力された単語は、最初の1文字を入力し変換キーを押すと、候補に現れるという方式も実用化されている。

【0003】 また、統計的な情報を用いる技術も知られている（「かな・漢字の文字連鎖情報を利用した統計的言語モデル」、山田智一ほか3名、日本音響学会講演論文集（平成3年3月）、95-96頁参照）が、ここで使用しているのは文字単位の情報であり単語としての扱いはない。

【0004】

【発明が解決しようとする課題】 ワープロなどで文章を入力する場合、一字一句正しい仮名を入力する必要がある。この負荷を軽減する方式として、上記のようなかな漢字変換時に既に文章中で入力された単語を変換候補の上位にするという方法や、既に入力された単語は、最初の1文字を入力し変換キーを押すと、候補に表示するという方式では、既に同じ単語が入力されていることが最低の条件である。しかし、全く同じ単語が複数回出現する確率は低い。

【0005】 一方、文書には内容によって、意味的につながりが深く、極めて出現しやすいパターンがある。たとえば、「拝啓、貴社 ますます」と入力されていれば、次にくる単語は「ご清栄の」とか「ご清祥の」とかいったものであることが多い。また、「長期研究計画」という言葉があれば、「立案／推進／見直し／提出」な

どの言葉が比較的良好に出現すると考えられる。

【0006】 そこで、既に入力された言葉の持つ意味的な情報を用いて、次に出現する単語を予測／表示することで、入力の負荷を軽減することが期待できる。

【0007】 本発明の目的は、既に入力された言葉の持つ意味的な情報を用いて、次に出現する単語を予測／表示し、入力の付加を軽減する機能を持った文章入力装置を提供することにある。

【0008】

【課題を解決するための手段】 本発明の文章入力装置は、文章を入力する入力手段と、入力された文章を格納する文章バッファと、文章バッファに蓄えられた文章から文章の意味ベクトルを作成する意味ベクトル作成手段と、文章の意味ベクトルから次の単語を予測する単語予測手段と、単語を予測するために記憶する意味辞書と、予測された単語を表示する予測単語表示手段と、予測された複数の単語から一つを選択する単語選択手段とを含む。

【0009】 上記単語予測手段は、文章の意味ベクトルと意味辞書に記述された単語の意味ベクトルとの類似度を比較して次の単語を選別する機能を具備する。

【0010】 さらに、上記単語予測手段は、単語を予測した後に入力された文字の情報をを用いて予測単語を絞り込む機能を具備する。

【0011】

【作用】 ユーザが、ある程度文章を入力すると、その後は、システムが単語を予測し、予測結果を表示する。ユーザの入力したい単語が表示された予測単語の中に存在すれば、ユーザは、キーやマウスなどの選択手法を用いて単語を選択する。予測単語中になければ、通常のワープロと同様にキーで入力を行う。予測単語にある場合には、単語を選択するだけで入力されるため、入力の負荷が低減される。また、予測を行うことで、意味的につながりのある単語が表示されるので、ユーザが文章を作成するときの補助にもなる。

【0012】

【実施例】 以下に本発明の文章入力装置について説明する。

【0013】 図1に、本発明の文章入力装置の実施例のブロック図を示す。

【0014】 図1において、1は、単語もしくは文章の入力部である。例えばキーボードやペン、タブレットと、文字認識装置、音声認識装置などに相当する。2は、入力された文章を格納する文章バッファである。通常、このバッファの内容の一部が表示装置に表示されるがここでは省略されている。3は、文章バッファの内容から入力された文章の意味特徴を作成する意味ベクトル作成部である。ここで作成される文章の意味ベクトルは、一つである必要はなく、複数の意味特徴が作成される可能性もある。4は、3で作成された文章の意味ベク

トルを用いて、その文章の意味ベクトルと関係の深い単語（もしくは単語列、文節、文章など）を予測する単語予測部である。単語予測部は、文章の意味ベクトルと後述の意味辞書に記述された単語の意味ベクトルとの類似度を比較して次の単語を選別する機能を具備している。

5は、3と4のブロックで使用される意味辞書である。

6は、4で予測された単語を表示する予測単語表示部である。7は表示された単語から、ユーザが単語を選択する単語選択部である。ユーザが入力したい単語がこの表示にない場合には、この選択は行われずに、通常のように（キーボードなどから）単語が入力される。また、この選択部で選択された語は、1の入力部から入力されたものと同様に文章バッファに格納され、文章の一部となる。

【0015】次に「長期研究計画の」という文章が入力されている場合の単語予測を例にあげて説明する。

【0016】まず、図2に示した意味辞書の例を説明する。ここでは、代表として9個の単語を示している。ここでの辞書は、それぞれの単語が横方向に6個の要素からなるベクトルで表現されている。意味辞書の作成は、各要素に意味づけを行い、その要素との関係を人手で付与しても作成可能であり、また、同一出願人による特許出願（特願平5-181265号）に記載のように、大量のデータから自動的に作成することも可能である。

【0017】この例では既に「長期研究計画」という言葉が入力されているので、この言葉を用いて文章の意味ベクトルを作成する。

【0018】図3に「長期研究計画の」という文章が含まれる場合の文章の意味ベクトルの計算例を示す。

【0019】図2の意味辞書から、「長期研究計画」という言葉を構成する各単語のベクトルは、長期（8, 1, 2, 4, 3, 2）、研究（7, 9, 4, 7, 2, 1）、計画（2, 5, 7, 3, 1, 1）であり、既に入力されたこれらベクトルを平均化することで文章の意味ベクトルが計算される（ただし、小数点以下は四捨五入）。即ち、 $(8+7+2) \div 3 = 5$ 、 6 、 $(1+9+5) \div 3 = 5$ 、 $(2+4+7) \div 3 = 4$ 、 3 、 $(4+7+3) \div 3 = 4$ 、 6 、 $(3+2+1) \div 3 = 2$ 、 $(2+1+1) \div 3 = 1$ 、 3 。この計算結果、これら3つのベクトルの平均（6, 5, 4, 5, 2, 1）が文章の意味ベクトルとなる。

【0020】ここでは、平均を取って文章の意味ベクトルとしたが、文章の意味ベクトルの作成方法は、荷重平均など各種の方法が考えられる。また、文章全体の意味ベクトルと、近傍（現在の入力に近い単語）だけを考慮した文章の意味ベクトルを両方作成するなど、複数作成することも考えられる。

【0021】次に、作成された文章の意味ベクトルから次の単語を予測する。予測に際しては、算出された文章の意味ベクトルと、意味辞書に記述された各単語の意味

ベクトルとの類似度を算出して、類似度が大きい（距離が小さい）ものから幾つか選択することで実現できる。選択する単語の数は、類似度の大きいものから順に固定した数にすることが考えられる。また、（対象とする単語の類似度／最大の類似度）が、一定のしきい値以上のものを選ぶことも考えられる。なお、最大の類似度とは、各単語の類似度のうち最大の値を指す。

【0022】文章の意味ベクトルと単語の意味ベクトルとの類似度は、ユークリッド距離（2乗誤差）を用いて算出することができ、図4に示すように、文章の意味ベクトル（6, 5, 4, 5, 2, 1）との距離が近いものの順に立案（距離：31）、推進（同：31）、見直し（40）、スイカ（74）、理科（89）、りす（107）となる（既に入力文章にある単語は省略した）。

【0023】この中で類似度順に三つの単語を表示する（ここでは、類似度の大きい順に三つ表示するように設定されている）と、立案、推進、見直しの単語が候補として表示される。ユーザは、この3つの単語を入力したければ、それらの候補をキーやマウスを用いて選択する。これらの単語以外であれば、従来のようにかな漢字変換や、文字認識などの手法を用いて入力を行う。

【0024】図10に、本発明の文章入力装置の操作手順のフローを示す。

【0025】1では、文章の初まりかどうかをチェックする。文章毎に話題が異なることを仮定しており、文章の初めであれば、2では文章ベクトルを初期化する。これは、文章の話題を示す意味ベクトルの要素をすべて0にすることで実現できる。3では単語が入力されたかどうかをチェックする。単語が入力され、確定されると、4ではその単語の意味ベクトルを意味辞書から抽出する。5では、今の単語の意味ベクトルを、既に入力された単語の意味ベクトルに基づき求められ抽出した文章の意味ベクトル（文章ベクトル）を用いて新たな文章ベクトルを作成する。つまり、3の単語入力確定すると、4及び5で文章ベクトルが更新される。6では、単語予測キーが入力されたかどうかをチェックする。予測キーが入力された場合には、5で得られた文章ベクトルと、意味辞書に格納されている単語の意味ベクトルとの類似度を7で計算する。8では、7で得られた結果を大きい順にソートする。9では、類似度の大きい順に幾つかを予測単語として表示する。10では、表示された単語の中からユーザが候補を選択したかどうかをチェックする。表示された候補が選ばれたなら、11でその単語を入力文字として確定する。候補選択以外のキーが入力されたなら、候補単語表示を消去し、次の入力待ちになる。

【0026】以上の単語予測の例では、入力された文章の意味ベクトルと近い意味ベクトルを持つ単語を表示する手法を示した。しかし、通常の文章入力では、1万語以上の単語が使用される。予測の精度が十分高ければ問

題がないが、予測精度が高くない場合には、従来の予測しない手法の場合の負荷とあまり変わらない。

【0027】次に、この点を改善するための手法を説明する。先の例では、「長期研究計画の」という文章が確定した場合には、3つの単語が予測され表示される。この改善手法では、次の1文字が入力された場合に、その1文字を用いて予測候補の限定を行う。つまり、なにも入力のない状態では先の3つの単語（立案／推進／見直し）が予測結果として表示されているが、次の1文字に「り」が入力されたとなると、その文字で始まるものの単語に限定される。図2の例では、距離が小さく、かつ「り」で始まる単語は、「立案」しかないので、「立案」だけが候補表示として残る。1文字単位ではなくて、ローマ字入力であれば、立案の「r」が入力された段階であっても、もしくは、「りつ」というように2文字以上が入力された段階でもそれぞれの情報を用いて制限をかけることが可能である。

【0028】もし、限定のない段階で、距離の小さい単語が非常に多く存在しても、次の入力を用いて候補を限定することで、最小限の入力を行い、その後は予測結果の単語を用いることが可能となり、入力の負荷が低減される。

【0029】この場合の実施例のブロック図を図5に示す。

【0030】図5において、入力部1、文章バッファ2、意味ベクトル作成部3、単語予測部4、意味辞書5、予測単語表示部6、及び言語選択部7のそれぞれは、図1に示したブロック図におけると同じ機能を持っているが、単語予測部4は、それ以外に、単語を予測した後に入力された文字の情報を用いて予測単語を絞り込む機能を具備している。したがって、意味ベクトル作成部3で作成された文章の意味ベクトルを用いて、その文章の意味ベクトルと関係の深い単語を単語予測部4で予測した後に、この予測単語を絞り込むために、次の文字を入力することを示している。

【0031】図6に、意味辞書を作成する手段のブロック図を示す。

【0032】図6において、1は、予め入力された文書を記憶する文章記憶部、2は、文章記憶部1に記憶された文書の形態素解析を行う文解析部、3は、文解析部2の形態素解析を行うために使用される文解析辞書、4は、文解析部2で解析された結果の自立語を記憶する自立語記憶部、5は、自立語記憶部4に記憶された自立語列と、後述する共起登録用辞書を用いて共起関係表を作成する共起関係作成部、6は、共起辞書を作成するときにどの自立語を用いるかを記述している共起登録用辞書、7は、共起関係作成部で作成された共起関係表（共起辞書）を格納する共起辞書記憶部である。詳細は同一出願人による特許出願（特願平5-181265号）に記載されている。

【0033】図7に、意味辞書の作成を説明する概念図を示す。

【0034】いま、入力された文章が、「業界の再編成の目標は合併によって経営基盤を拡充し、過剰な設備を削減することだ。」（図7a）とする。入力された文章は文解析部2で解析され、自立語「業界、編成、目標、合併、経営、基盤、拡充、過剰、設備、削減」（図7b）が抽出される。抽出された自立語は自立語記憶部4に記憶される。一文が入力し終わると、自立語記憶部4に記憶された自立語によって共起関係が作成される。図7の入力例文では、その下に記載された抽出自立語が得られる。この抽出自立語から共起関係表（＝意味辞書）を作成する（図7c）。この場合、抽出自立語から重複を許して二つの自立語を取り出し、その二つの自立語Aと自立語Bとする。共起関係表には、A-Bの組み合わせに相当する昇目に1を加える。つまり、同じ文章に現れた言葉の組み合わせは、なんらかの関係を持つことと仮定している。

【0035】例えば、「業界」と「編成」が選択されたならば、共起関係の横軸（行方向）が「業界」で、縦軸（列方向）が「編成」である昇目に1を加える（図7cにおいて1が記載されている）。抽出された自立語の中から全ての二つの自立語の組を抽出して、該当する共起関係表の要素に1を加えることで共起関係表が作成される。このようにして、意味辞書が作成される。

【0036】図8及び図9に、研究計画審議資料を作成するための文章を作成する例を示す。

【0037】図8では、「今年度の研究計画について、開発」まで入力した段階で、「単語予測」を行うためのキーを押すと、その下の部分の単語予測結果が表示される。現在、その予測の第1位が「日程」であるため、「日程」は、入力されるべき単語の候補として、本文中にも表示されている。もし、次候補のキーを押すと、次の予測候補である「予定」が「日程」と変わって表示される。取り消しキーが押されたなら、予測結果がすべて消えて、通常の入力と同じになる。取り消し、次候補以外のキーが入力されたなら、今表示されている予測単語を確定し、予測結果は消去される。なお、「単語予測」や「次候補」「予測取り消し」などのキーは独立してある必要はなく、かな漢字変換の変換キーや取り消しキーと併用してもかまわない。

【0038】図9では、図8と同じく、「今年度の研究計画について、開発」まで入力され、さらに「も」という一文字が入力された場合を示している。「も」が入力されてから「単語予測」キーを押すと、「も」で始まる単語のみが予測結果として表示される。これにより、より正確な予測を行うことができる。勿論、単語予測キーを押さなくても、予測結果を出すことも考えられる。この場合、画面のどこかに、常時現在の予測候補が表示されている。文字が入力されるたびに、その情報を用いて

7

予測候補が更新される。ユーザは、もし、候補の中に入力したい単語があれば、それを選択することで、単語の入力が可能となる。勿論、単語ではなく、例文を予測候補として出すことも考えられる。

【0039】

【効果】本発明により、既に入力された文章から次に入力される単語もしくは文節、文章を予測し、それをユーザに選択させることで、全ての文章を入力することなく、選択操作のみだけで単語を入力でき、入力負荷を減らすことが可能となる。

【図面の簡単な説明】

【図1】本発明の実施例のブロック図である。

【図2】意味辞書の例を示す図である。

【図3】文章の意味ベクトルの計算例を示す図である。

【図4】意味ベクトルによる距離を示す図である。

【図5】本発明の別の実施例のブロック図である。

8

【図6】意味辞書を作成する手段のブロック図である。

【図7】意味辞書の作成を説明する概念図である。

【図8】本発明の実施例による表示例を示す図である。

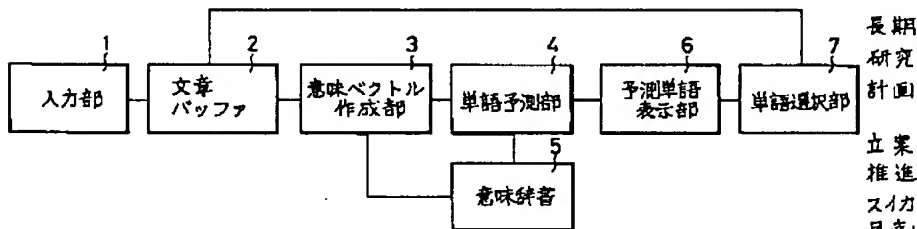
【図9】本発明の別の実施例による表示例を示す図である。

【図10】本発明の文章入力装置の操作手順のフローを説明する図である。

【符号の説明】

- 1 入力部
2 文章バッファ
3 意味ベクトル作成部
4 単語予測部
5 意味辞書
6 予測単語表示部
7 単語選択部

【図1】



【図2】

長期研究計画	8	1	2	4	3	2
	7	9	4	7	2	1
	2	5	7	3	1	1
立案推進スイカ見直りす理科	3	5	7	2	2	3
	2	4	6	2	1	1
	4	1	2	1	5	6
	3	5	9	4	2	3
	2	4	1	2	8	2
	4	1	2	1	2	8

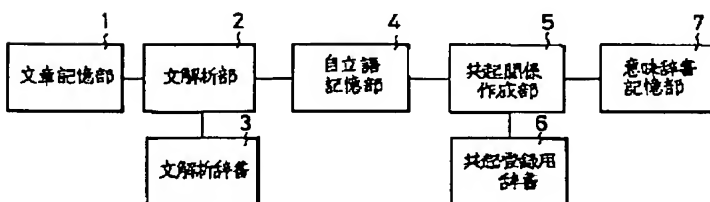
【図3】

長期 (8, 1, 2, 4, 3, 2)
研究 (7, 9, 4, 7, 2, 1)
計画 (2, 5, 7, 3, 1, 1) } 平均 (6, 5, 4, 5, 2, 1)

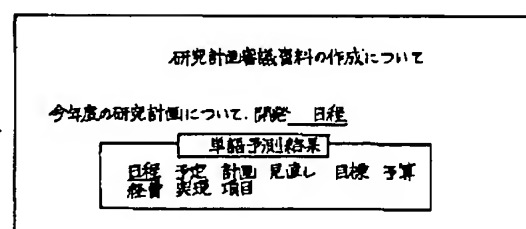
【図4】

文章の意味ベクトル	6	5	4	5	2	1
立案推進スイカ見直りす理科	3	5	7	2	2	3
	2	4	6	2	1	1
	4	1	2	1	5	6
	3	5	9	4	2	3
	2	4	1	2	8	2
	4	1	2	1	2	8
距離	3	1				
	3	1				
	7	4				
	4	0				
	1	0	7			
	8	9				

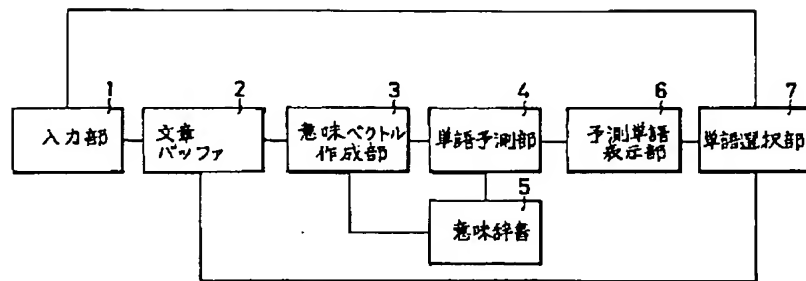
【図6】



【図8】



【図5】



【図7】

(a)

業界の再編成の目標は合併によって経営基盤を拡充し、過剰な設備を削減することだ。

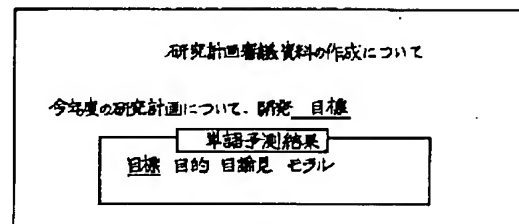
(b)

業界	編成	目標	合併	経営	基盤
拡充	過剰	設備	削減		

(c)

	業界	編成	目標
業界	1		1 1
編成	1	1 1	
目標	1	1 1	

【図9】



【図10】

